



# Modelli, campo e prospettive della Digital History. Scenari, contesti e applicazioni

Roberto Testa  
Dipartimento di Studi Storici  
Università degli Studi di Torino

## Abstract

L'obiettivo di questo contributo è discutere alcune delle caratteristiche del campo della Digital History, nel più generale contesto delle trasformazioni che stanno investendo i modelli di organizzazione della conoscenza e le pratiche disciplinari ad esse riferite. Nell'ambito specifico della cultura storica la diffusione sempre più pervasiva delle culture digitali ha condotto alla profilazione, ancora incerta, della Digital History. Nella fase attuale, che seguendo Luciano Floridi possiamo definire della "Iperstoria", quali sono principi, obiettivi, metodi della Digital History? Quali sono i risultati ottenuti dai processi di sperimentazione in corso? E infine, il potere computazionale delle macchine sta modificando alcuni elementi fondativi della metodologia della ricerca storica? L'articolo propone una ricognizione generale su questi argomenti, cercando di integrare in un ragionamento unitario questioni che toccano elementi decisamente centrali all'interno di una disciplina fondativa della tradizione culturale.

## Models, Field and Perspectives of Digital History. Scenarios, Contexts and Applications

This essay aims to discuss some of the characteristics in the field of Digital History. My analysis focuses on the general context of the transformations that are affecting the models of knowledge organization and the associated disciplinary practices. Which are the principles, objectives and methods of Digital History in this historical phase that Luciano Floridi defines "Hyperhistory"? Is the computational power of machines actually modifying some of the fundamental elements of historical research methodology? This article intends to deliver a general analysis on these topics, while also trying to integrate in a unitary reasoning various issues that touch on elements of extreme importance in a discipline that is recognized as one of the cores of cultural tradition.

*Published 2 May 2021*

Correspondence should be addressed to Roberto Testa, Dipartimento di Studi Storici, Università degli Studi di Torino, via S. Ottavio 20, Torino. Email: [roberto.testa.9@virgilio.it](mailto:roberto.testa.9@virgilio.it)

*DigitCult, Scientific Journal on Digital Cultures* is an academic journal of international scope, peer-reviewed and open access, aiming to value international research and to present current debate on digital culture, technological innovation and social change. ISSN: 2531-5994. URL: <http://www.digitcult.it>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (IT) Licence, version 3.0. For details please see <http://creativecommons.org/licenses/by/3.0/it/>



## Nota Introduttiva

La diffusione del digitale ha trasformato – e continua a trasformare – in tutti i sensi i modelli di organizzazione della conoscenza che a loro volta modificano principi, metodi ed applicazioni dei diversi campi disciplinari del sapere. Il progresso della tecnologia, principalmente basato sulla digitalizzazione della comunicazione e sulle reti informatiche, ha per la prima volta ricombinato le informazioni e le forme di comunicazione all'interno di un contesto digitale interattivo entro cui si ritrovano le eterogenee culture veicolate dall'essere umano (Castells 2014). Nella transizione dall'analogico al digitale gli strumenti (algoritmi, software) intrinsecamente si avvicinano sempre più alla natura delle loro risorse (i dati e le fonti), fino al punto da divenire quasi omogenei, della stessa sostanza (Floridi 2017). In questo scenario si sviluppa l'ampia e discussa area delle Digital Humanities, quindi un differente modo di rappresentare e accedere alla cultura umanistica (Numerico et al. 2010). Lavorare e ragionare all'interno del campo delle Digital Humanities significa mettere insieme conoscenze eterogenee con lo scopo di analizzare e presentare forme di sapere e allo stesso tempo utilizzare tecnologie innovative per fare ricerca. Questa assume una nuova visione che è rappresentata dalla computabilità del dato (Tomasi 2019), che diventa un elemento dirimente con cui lo storico, e in generale lo studioso, deve inevitabilmente confrontarsi, in quanto talvolta rappresenta la fonte e il punto di partenza della sua euristica. Per riprendere Marc Bloch: “[...] lo strumento non fa la scienza. Ma una società che abbia la pretesa di rispettare le scienze non dovrebbe disinteressarsi dei loro strumenti” (Bloch 2009, 117).

Da qui nasce l'esigenza di riassumere un insieme di pratiche, di strumenti e di esperienze sotto l'ombrello della Digital History, a sua volta situata all'interno dell'ambito delle Digital Humanities. Infatti, con il passare del tempo, la disciplina tradizionalmente intesa subisce l'influenza del digitale, trasformandosi ed effettuando particolari parabole per superare tutti gli ostacoli che incontra nel suo cammino.

Nella più vicina contemporaneità la Storia diventa chiave di lettura per interpretare l'età dell'informazione e la società, e, secondo Luciano Floridi, diventa “iperstoria” dal momento in cui fa affidamento sulle ICT (Information and Communication Technology) per registrare, collegare, utilizzare dati, informazioni di ogni genere (Floridi 2017). Più precisamente la realtà informativa entro cui ci muoviamo è definibile infosfera: all'interno di questo ambiente informazionale si ritrovano insieme mondo analogico e digitale, in una fusione che può tuttavia generare sia confusione sia sincronizzazione di orizzonti e spazi (Vivarelli 2021).

L'obiettivo del presente contributo è indagare l'ambito della Digital History all'interno del contesto metodologico della ricerca storica, analizzando strumenti, fonti e modelli che permettano di maturare riflessioni utili in prospettiva futura sia da un punto di vista teorico sia da uno pratico.

Nel primo paragrafo si darà conto della discussione riguardo le diverse definizioni di Digital History, Digital Public History e Digital Historian, per poi delineare una sintesi dell'evoluzione del rapporto tra cultura storica e cultura digitale, riprendendo lo stato dell'arte e parte del lavoro svolto negli ultimi quaranta anni da storici e informatici. Successivamente verranno presentati in rassegna alcuni dei principali strumenti della ricerca storica e alcune prospettive legate all'applicazione della Intelligenza Artificiale. In un successivo paragrafo verranno presi in esame due progetti inerenti all'ambito della Digital History: *Slaves Voyages* e *ALCIDE*. Il primo è un database non testuale sulle tratte schiavistiche globali (tra il XVI e il XIX secolo) che offre varie possibilità di visualizzazione e interrogazione del dato. Il secondo è un database testuale che raccoglie gli scritti dello statista Alcide De Gasperi permettendone una particolare analisi e ricerca grazie a specifici *tools*. Entrambi i progetti verranno poi criticamente discussi, nel secondo caso grazie a un confronto diretto con sviluppatori e membri del *team* di ricerca. Il quarto paragrafo proporrà, infine, delle riflessioni conclusive sul percorso della Digital History e sul peso che questo può assumere a livello teorico-metodologico all'interno dell'ambito disciplinare della ricerca storica.

In questa fase di transizione in cui la dimensione digitale assume una sempre più rilevante influenza all'interno degli studi umanistici, lo storico si trova davanti a un bivio: procedere secondo la metodologia tradizionale o seguire – ma anche tracciare – i nuovi percorsi digitali sfruttando il potere computazionale delle macchine. Qualcuno sceglie la prima via, altri la seconda, altri ancora cercano una terza possibilità – che oggi, agli occhi di chi scrive, pare la più concreta – rappresentata dalla mediazione e dall'integrazione tra entrambe le opzioni,

sfruttando ora la potenza degli strumenti tecnologici, ora il profondo sguardo interpretativo che solo grazie al confronto critico con la complessità delle diverse tipologie di documenti, ed i loro molteplici contesti, può essere acquisita.

## Alcune definizioni: Digital History, Digital Public History, Digital Historian

L'ambiente della Digital History appartiene al mondo ormai ampiamente delineato delle Digital Humanities, che, per quanto la definizione sia parecchio discussa, può essere sommariamente definito come un

“Campo di ricerche interdisciplinari il cui oggetto è lo studio degli artefatti e dei processi culturali, tradizionale dominio delle scienze umane, nelle loro varie forme espressive [...] con metodologie, linguaggi e strumenti informatici.” (Lessico del XXI secolo, 2012)<sup>1</sup>.

Il termine viene introdotto per la prima volta da alcuni editori della Blackwell Publishing, che nel 2004 sostituirono il loro *Companion to Humanities Computing* in *Companion to Digital Humanities* per un'operazione di marketing che presta più attenzione al cambiamento introdotto negli studi umanistici dal digitale (Henrickson 2019). Le Digital Humanities si pongono l'obiettivo di trasformare il modo in cui i documenti e i fenomeni possano essere affrontati, trasmessi, interpretati, interrogati e immaginati dagli studi umanistici con l'influenza e il supporto del digitale (Nyhan and Flinn 2016, 1).

Far rispecchiare queste definizioni sull'ambiente della ricerca storica sicuramente crea non poche difficoltà e dà vita a continui dibattiti. I primi a utilizzare pubblicamente il termine Digital History furono Ed Ayers e William Thomas nel 1997, in seguito al progetto digitale “The Valley of The Shadow”, nato per l'insegnamento della storia e che avrebbe portato nel 1998 alla nascita della Virginia Center for Digital History (Nystrom and Nystrom 2005). Successivamente il «Journal of American History» definisce la Digital History come qualsiasi metodologia di ricerca, rivista, blog o attività che utilizzi tecnologie digitali nel lavoro di ricerca, nell'insegnamento e nella diffusione della storia (Cauvin 2016, 175). Se la prima definizione fa riferimento a una specifica progettualità, la seconda risulta essere troppo generica e semplicistica, oltre che imprecisa. Un'altra definizione è presentata da Seefeldt e Thomas nella rivista statunitense «Perspectives on History»:

“[...] An approach to examining and representing the past that works with the new communication technologies of the computer, the internet network, and software systems. [...] to make, define, query, and annotate associations in the human record of the past.” (Seefeldt and Thomas, 2009).

Un suggerimento giunge da Gerben Zaagsma che riprende una definizione di Digital Humanities proposta da Schnapps e Presner:

“[...] *an array of convergent practices* [enfasi di Zaagsma] that explore a universe in which: a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; [...] and b) digital tools, techniques, and media have altered the production and disseminations of knowledge in the arts, human and social sciences.” (Zaagsma 2013, 15).

Tutte le definizioni fino ad ora riproposte convergono su un punto: la Digital History non è considerata una sotto-disciplina della storia e non ha una sua autonomia all'interno dei rami del sapere. Di conseguenza non è ancora riuscita ad attecchire nella comunità scientifica e accademica, anche perché nasce come diretta risposta alle opportunità e ai problemi dovuti alla

<sup>1</sup> In questo caso si fa riferimento alla relativa voce 'Informatica umanistica' presente nel 'Lessico del XXI Secolo' (2012): [https://www.treccani.it/enciclopedia/informatica-umanistica\\_%28Lessico-del-XXI-Secolo%29/](https://www.treccani.it/enciclopedia/informatica-umanistica_%28Lessico-del-XXI-Secolo%29/).

diffusione del digitale all'interno della pratica storica, quindi può essere intesa piuttosto come un considerevole aggiornamento della metodologia di ricerca (Salvatori 2017, 62).

Questo nuovo approccio – così come tanti altri – talvolta non è visto di buon occhio e continua a suscitare discordanze in materia di metodologia all'interno della stessa comunità accademica, perlomeno in Italia. Tra i suoi principali sostenitori si trova Serge Noiret, presidente dell'AIPH, che definisce la Storia digitale come un utilizzo delle risorse digitali atto a produrre contenuti in rete e far sì che questi vengano preservati nel tempo (Noiret 2011, 209). A tal proposito Lemerrier ritiene il digitale ciò che ha trasformato il mestiere di storico, costringendolo ad aggiornarsi per ampliare le sue possibilità nel campo della visualizzazione dei documenti, nella creazione di bibliografie, nell'accessibilità alle fonti e in altre attività che possano facilitare la ricerca, la pubblicazione e il confronto tra studiosi stessi e tra studiosi e pubblico (Grandi and Ruiz 2012). Ed è proprio al pubblico che si fa riferimento: la storia digitale ha il compito di aprirsi a un pubblico più ampio, quindi di fare attenzione ai nuovi interlocutori, soprattutto se non accademici. Lo storico digitale, secondo la definizione di Enrica Salvatori

“Condivide – deve condividere – con gli storici ‘tradizionali’ il rigore e i metodi di analisi, che però contestualizza nel mondo digitale, verificandone l'applicabilità e valutandone i risultati al fine di proporre nuovi metodi e nuovi strumenti di analisi, complementari e non sostitutivi a quelli in uso [...]” (Salvatori 2017, 68).

La figura del Digital Historian viene inoltre definita da Anna Nilsson Hammar come quella di un produttore e consumatore di dati che può utilizzare i media digitali per presentare e diffondere informazioni a un pubblico in rete (Nilsson Hammar 2015, 101).

Anche in questo ultimo caso la definizione appare assai generalista, in quanto un produttore di dati, informazioni e conoscenza non deve obbligatoriamente essere uno storico; in particolare lo storico non può essere trasformato, come in questo caso, nella sola figura di divulgatore digitale, in quanto per sua definizione lo storico prima di essere un narratore è un ricercatore, uno scienziato. Considerare la tecnologia, i computer e internet esclusivamente come semplici strumenti dello storico appare oggi molto limitante, poiché, come si avrà modo di vedere nei paragrafi successivi, fonti, dati e informazioni vengono prodotti giornalmente dalla rete e da chi naviga.

## L'evoluzione della cultura storica e la Digital History

Il primo approccio al mondo dell'informatica da parte degli storici viene definito Historical Computing, quindi applicazioni informatiche adeguate alle esigenze della ricerca storica (Thaller 1991). Negli anni '70 storici economici, storici urbani e storici della politica utilizzavano quotidianamente le fonti quantitative elaborate dai computer, fino ad arrivare al punto di costruire *databases* contenenti una gigantesca mole di informazioni collegate tra loro (Noiret, 2008). Nel decennio successivo un esperimento interessante fu quello di Manfred Thaller, che elaborò una *workstation*, “Clio”,

“[...] basata su una complessa architettura modulare, in grado di offrire allo storico tutti gli strumenti di analisi di cui egli potesse aver bisogno, da quelli statistici e metrologici a quelli cronologici, cartografici o iconografici, integrabili tra loro attraverso il collegamento nominativo dei record (*record linkage*).” (Delle Donne 2005, 5).

Lo stesso Thaller si sarebbe reso portavoce, attraverso l'articolo *The Need for a Theory of Historical Computing*, della necessità di gettare alcune prime basi teoriche per quella che sarebbe stata – e che continua ad essere – la Digital History. Con l'avvento del WWW gli storici, che fino ad allora avevano condiviso parte dei loro materiali attraverso *mailing list*, si immerse nel suo *mare magnum* portando avanti diversi progetti (Nelson 1999). Uno dei primi progetti si materializzò nel 1991 grazie all'idea di Donald Mabry, che fondò RA, sito FTP basato sulla condivisione e il trasferimento diretto di *file* e dati (sistema *client-server*): il sito offriva materiale per storici all'interno di un archivio digitale (10 *megabyte* di memoria), oltre che gli indirizzi delle varie liste di discussioni accademiche internazionali che nel frattempo si erano venute a creare<sup>2</sup>.

<sup>2</sup> Grazie al racconto diretto dell'esperienza di Donald Mabry è possibile ricostruire la storia di RA e

A questo seguì HNSource, sviluppato nel 1993 da Lynn H. Nelson che, attraverso Lynx<sup>3</sup>, un *web-text browser*, permetteva agli utenti di leggere materiale più velocemente occupando meno CPU. Questi progetti, come altri, rappresentano la testimonianza della volontà degli storici di estendere il campo della ricerca verso il pubblico, oltre che rendere accessibili documenti e fonti per creare, riprendendo le parole dello stesso Nelson, “[...] a world in which everyone would have a great library at their fingertips.” (Nelson 1999).

Ma quale elemento ha mutato il modo di operare dello storico? La questione fondamentale è rappresentata da uno dei punti inscindibili della ricerca: le fonti. Quale cambiamento ha quindi introdotto il digitale nel rapporto tra storico e fonti?

La prima distinzione rilevante è tra fonti native digitali e tra fonti digitalizzate: le prime sono *digital born*, direttamente prodotte dal e sul digitale, mentre le seconde sono fisiche e trasposte in rete, quindi dematerializzate (Detti and Lauricella 2007). Le fonti native digitali corrispondono a una serie di numeri (con una logica che segue il codice binario) che viene letta dal sistema e poi rappresentata graficamente o testualmente, mentre le fonti digitalizzate possono presentarsi in altre forme alla vista umana (generalmente documenti scannerizzati tramite lettori ottici). Ulteriore differenza è rappresentata dalla possibilità offerta dai nuovi metalinguaggi di marcatura (SGML e XML) che si possono applicare sulle prime: questi permettono di inserire *tag* e *metadati* all'interno della fonte, in grado di descriverne struttura, articolazione e relazioni senza alterare la natura, l'identità e l'originalità del testo (Falletta 2010). Ecco allora che i documenti si arricchiscono di informazioni e aiutano lo storico nel lavoro di selezione, sintesi e comparazione, oltre che di presentazione. Ma la novità più interessante è costituita dall'esponenziale crescita di dati digitali, nuove fonti prodotte dallo stesso Web e utili alla ricerca se considerati nel loro insieme molto ampio (Big Data).

Per certi versi il lavoro che lo storico ha compiuto con le fonti tradizionali e materiali rimane tale anche con quelle digitali, quindi l'interrogazione, l'analisi e la lettura critica. A oggi purtroppo non si hanno dei veri e propri strumenti per affermare la validità di un documento digitale, se non quelli che lo storico dovrebbe possedere per via della sua stessa formazione: metodologia, rigore e sistematicità<sup>4</sup>. Del resto, come sosteneva Marc Bloch, “[...] una menzogna, in quanto tale, è, a suo modo, una testimonianza.” (Bloch 2009, 133).

Svelare le motivazioni per cui è stato pubblicato o diffuso qualcosa di falso, di rimaneggiato, di completamente inventato e spacciato come fonte originale, è un'operazione attraverso la quale si possono ottenere importanti informazioni prima nascoste. In quest'ottica può far riflettere il caso dello storico Orlando Figes, che nel 2010, dopo essere stato citato in giudizio, ha ammesso di aver pubblicato recensioni anonime (o sotto pseudonimo) su Amazon in cui criticava fortemente pubblicazioni e libri di alcuni suoi colleghi rivali sponsorizzando al contempo i propri lavori<sup>5</sup>.

La natura del digitale è in primis rappresentativa: una volta che si spegne un computer, che si chiude un programma, il contenuto scompare alla vista umana (Vitali 2004). Tuttavia, una questione aperta rimane quella della conservazione dei file e delle pagine web: se da un lato il progresso tecnologico amplia le possibilità di lavoro, dall'altro lato rende sempre più necessaria la continua conversione di file o la trasformazione dei dispositivi di lettura.

Per quanto riguarda le pagine web, in parte risponde a questa esigenza la Wayback Machine, iniziativa di Archive.org che dal 2001 ha il compito di archiviare e conservare i siti web che non sono più attivi (circa 424 miliardi, fonte <https://archive.org/web>). Il dispositivo permette a ogni utente di “salvare” in un qualsiasi momento una pagina di un sito web per poterlo citare in futuro e ritrovarlo. Così facendo, la macchina ripercorre la storia – attraverso le captures, istantanee – di ciascun sito web archiviato, permettendo all'utente di muoversi al suo interno, selezionando una specifica data per recuperare informazioni utili o per vedere le modifiche che sono state attuate in un determinato arco di tempo<sup>6</sup>.

---

successivamente di Historical Text Archive, dominio web realizzato dallo stesso Mabry, divenuto nel frattempo Direttore dell'Institute for the Humanities alla Mississippi State University, <http://www.historicaltextarchive.org/about.php>.

<sup>3</sup> Lynx, Lynx Information, <https://lynx.brower.org/>.

<sup>4</sup> Per un ulteriore approfondimento sull'autenticità si veda Tommaso Detti and Giuseppe Lauricella, “Una storia piatta? Il digitale, Internet e il mestiere di storico.”, *Contemporanea*, 10.1 (2007): 3-23.

<sup>5</sup> Orlando Figes è uno storico contemporaneista studioso della Russia, professore presso il Birkbeck College di Londra. Si può approfondire la vicenda delle *fake reviews* su <https://www.bbc.com/news/uk-10670407>; <https://www.theguardian.com/books/2010/apr/23/historian-orlando-figes-amazon-reviews-rivals>.

<sup>6</sup> Si può prendere come esempio il caso di Digital History, un portale americano di storia digitale, in

## Strumenti a disposizione del Digital Historian

Un elemento per la ricerca di documenti, testi e fonti è la banca dati o *database*, contenitore elettronico di dati che possono essere interrogati in vario modo. Il produttore del *database* per aiutare il visitatore a orientarsi può fornire vari *tools* come la ricerca per parole chiave, per soggetto, per autore, per anno di pubblicazione. Uno dei più noti *database* in ambito scientifico è JSTOR<sup>7</sup>, che contiene circa 12 milioni di testi (tra libri, giornali, riviste, collezioni tematiche, fonti primarie); tra gli altri si ricordano anche Google Scholar<sup>8</sup>, ProQuest<sup>9</sup> e LexisNexis<sup>10</sup>.

Altra risorsa molto utile è rappresentata da Zotero<sup>11</sup>, *tool Open Source* creato per organizzare, citare, catalogare e condividere la ricerca bibliografica. A tal proposito non si possono non menzionare i vari OPAC (Online Public Access Catalogue), cataloghi ad accesso aperto che permettono di effettuare ricerche sui vari testi delle biblioteche secondo diversi criteri e visualizzare le informazioni bibliografiche a riguardo<sup>12</sup>.

Con l'introduzione dei Big Data, raccolte di *dataset* troppo grandi e variegati per essere processati singolarmente e manualmente (Cellini 2018), lo storico ha a disposizione nuovi strumenti e nuove fonti per il proprio lavoro. Si passa così dal concetto di fonte a quello di dato, inteso come un fenomeno espresso in un formato quantificato (*datafication*, datificazione) in modo da poter essere registrato, analizzato e riorganizzato (Mayer-Schoenberg and Cukier, 2013). I Big Data possono essere prodotti dall'interazione che l'uomo ha con il computer o generati automaticamente dalle macchine, raccolti dai sensori (come il GPS) o prodotti dall'utilizzo dei *social network*, *websites*, *chat* (Hurwitz et al. 2013). Per quanto riguarda la ricerca storica il loro utilizzo può rispondere ad alcune intrinseche difficoltà che il metodo tradizionale incontra davanti a *corpus* di fonti di dimensioni spropositate, motivo per cui si discute spesso sulla sovrabbondanza di risorse a disposizione dello storico (Pancur et al. 2016). I Big Data offrono allo storico uno sguardo macroscopico comprimendo via via la complessità per risalire alle strutture e alle relazioni nascoste, rendendole chiare alla vista, in forte contrapposizione a quella visione di partenza microscopica tipica della metodologia tradizionale (van Eijanetten et al. 2013).

"The rise of the "digital" pushes historians into new areas of research and away from micro histories centering around well understood archives, texts, datasets and linked to relatively short time scales and played out in relatively small places."  
(François et al. 2016).

Data la continua crescita e l'enorme quantità di dati di cui si dispone – al punto da dover ricorrere all'unità di misura dello *zettabyte*<sup>13</sup> – per lo storico è fondamentale sapere cosa si sta cercando, onde evitare la dispersione e la confusione generate dalla sovrabbondanza di informazioni.

Uno degli scopi del lavoro sui Big Data – e, in generale, secondo chi scrive, quello della Digital History – non è quello di sostituire l'ermeneutica e l'interpretazione propria del lavoro dello storico ma quello di affiancare e integrare entrambi gli approcci nel lavoro di ricerca:

"[...] computational systems create data from narrative, while historians create narratives from data. The underlying assumption is that knowledge is derived from information which is derived from data, and there the process must begin. [...]"

---

<https://web.archive.org/web/20050901032007/http://www.digitalhistory.uh.edu/>.

<sup>7</sup> <https://www.jstor.org>.

<sup>8</sup> <https://scholar.google.com>.

<sup>9</sup> <https://www.proquest.com/researchers/student-general-researcher/>.

<sup>10</sup> <https://www.lexisnexis.com/en-us/gateway.page/>.

<sup>11</sup> <https://www.zotero.org>.

<sup>12</sup> Il caso più noto nel panorama italiano è sicuramente l'OPAC SBN, Catalogo del Servizio Bibliotecario Nazionale, <https://opac.sbn.it/opacsbn/opac/iccu/free.jsp>.

<sup>13</sup> Floridi, a tal proposito, riprende la definizione di *età dello zettabyte* per descrivere la nostra contemporaneità (Floridi 2017, 47). Di seguito la definizione di *zettabyte* secondo TechTerms: "A zettabyte is 10<sup>21</sup> or 1,000,000,000,000,000,000 bytes. One zettabyte (abbreviated "ZB") is equal to 1,000 exabytes and precedes the yottabyte unit of measurement. [...] Because the zettabyte unit of measurement is so large, it is only used to measure large aggregate amounts of data. Even all the data in the world is estimated to be only a few zettabytes." <https://techterms.com/definition/zettabyte>

knowledge can ever create information, which can be manifested in data.” (Bozic et al. 2016, 96).

I Big Data in questo senso rappresentano un elemento dirimente, di discontinuità, all'interno dell'evoluzione della disciplina della ricerca storica. Da questo momento in poi, a patto che non voglia intenzionalmente ignorare le nuove fonti, lo storico non può più procedere come ha sempre fatto. Questo è ciò che distingue il Digital Historian dallo storico che semplicemente utilizza strumenti digitali nel suo lavoro, quindi secondo chi scrive si può cautamente parlare di Digital History.

Un interessante esempio – in corso di realizzazione – è rappresentato da Time Machine Europe<sup>14</sup>, rete continentale che ha lo scopo di lavorare con i *Big Data* del passato, ricavati dalla digitalizzazione del patrimonio archivistico e museale, per mantenere viva la storia culturale, geografica e sociale dell'Europa. Come si evince dal *Manifesto*, Time Machine intende contribuire allo sviluppo economico a livello di ICT (quindi con applicazioni di *Virtual Reality* e *Augmented Reality*) per sostenere le industrie creative e il turismo, oltre che lo sviluppo di *smart cities, urban planning e land management*<sup>15</sup>.

Il lavoro della rete Time Machine si concretizza attualmente in varie *Local Time Machines*, “macchine del tempo”, progetti partecipativi che si pongono come obiettivo la ricostruzione in 4D di aree o città utilizzando diversi *tools*. Un progetto pioniero è quello della città di Venezia: la *Time Machine* ricopre un arco temporale che va dall'anno 1000 al 2000 grazie alla digitalizzazione attraverso software altamente specializzati di migliaia di documenti (oltre 80km), manoscritti, carte geografiche e opere d'arte<sup>16</sup>; i documenti possono inoltre essere analizzati con i *tools* di Text Analysis per scoprirne relazioni semantiche e tra lemmi.

Per Data Analysis, infine, ci si riferisce al processo di applicazione sistematica di tecniche logiche e/o statistiche per descrivere, illustrare, riassumere e valutare i dati: la sua metodologia cambia in base all'ambito di studio, all'approccio che si assume nei confronti dell'argomento (ad esempio la storia orale, quindi l'intervista) e dalla forma dei dati (note di campo, documenti, riprese video o audio)<sup>17</sup>. Riferendosi ai Big Data, le sue possibilità di approccio sono varie e non tutte hanno la stessa utilità per lo storico; vanno annoverate, tra le più utili, il Data Mining e la Data Visualization.

Il Data Mining, dallo stesso significato del termine (“estrazione dei dati”), è il processo che permette, mediante l'applicazione di algoritmi che individuano associazioni nascoste, l'estrazione di elementi significativi all'interno di *dataset* che contengono dati talvolta di natura eterogenea (Guerrini and Possemato 2015). La macchina ha bisogno di tradurre il linguaggio umano in dati per poterlo analizzare e comprendere mediante algoritmi. Il Data Mining è una delle tecniche di KDD (Knowledge Discovery in Databases), “[...] processo di identificazione di validi, nuovi, potenzialmente utili e in definitiva comprensibili schemi (*patterns*) nei dati.” (*trad.* da Ghos 2008, 3). Un'altra forma di approccio è costituita dalla Semantic Text Analysis, che può essere applicata per ottenere più informazioni sull'argomento specifico dalla grande mole di dati testuali senza averli letti, prestando particolare attenzione sul contenuto storico e contestuale delle parole e delle frasi contenute all'interno del *dataset* (van Eijatten et al. 2013).

La Data Visualization è la rappresentazione grafica dei dati, che deriva da un'esplorazione visuale di dati di qualsiasi dimensione. Il procedimento di visualizzazione permette allo studioso

<sup>14</sup> Si tratta di una rete nata nel 2019 e che comprende più di 14.000 istituzioni e oltre 100.000 professionisti tra scienziati, informatici, storici, ricercatori, umanisti, archivisti, etc., con sede centrale a Vienna e sedi anche a Losanna, Amsterdam e Budapest. <https://www.timemachine.eu/about-us/>.

<sup>15</sup> Time Machine Manifesto. <https://www.timemachine.eu/wp-content/uploads/2019/06/Time-Machine-Manifesto.pdf>.

<sup>16</sup> Il progetto viene portato avanti dall'ingegnere informatico Frédéric Kaplan e dalla storica Isabella di Lenardo in collaborazione con l'Università Politecnica di Losanna (EPFL) e con l'Università Ca' Foscari di Venezia. Per ulteriori approfondimenti si veda Alison Abbott. “The ‘time machine’ reconstructing ancient Venice’s social networks.” *Nature*, 14 giugno 2017 (<https://www.nature.com/news/the-time-machine-reconstructing-ancient-venice-s-social-networks-1.22147>). Nel 2019 il progetto ha subito una brusca interruzione dovuta a una discussione tra l'Archivio di Stato di Venezia e l'EPFL. Per un approfondimento a riguardo si veda Davide Castelvechi. “Venice ‘time machine’ project suspended amid data row.” *Nature*, 25 ottobre 2019, (<https://www.nature.com/articles/d41586-019-03240-w>). Un video di presentazione del progetto è disponibile su YouTube, “A virtual time machine for Venice” (<https://www.youtube.com/watch?v=uQQGgYPRWfs>).

<sup>17</sup> Per un approfondimento si veda Responsible Conduct of Research, *Data Analysis*, in [https://ori.hhs.gov/education/products/n\\_illinois\\_u/datamanagement/datopic.html/](https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/datopic.html/).

di poter vedere rappresentato graficamente l'oggetto della propria analisi. Andy Kirk la definisce "[...] the representation and presentation of data to facilitate understanding." (Kirk 2016, 31).

La Data Visualization si è evoluta fino a permettere a ricercatori e utenti di avere una panoramica visivamente più chiara e compatta di un insieme di dati che altrimenti sembrerebbero tra loro disconnessi.

Gli sviluppi e le prospettive future della ricerca storica guardano anche verso le nuove sperimentazioni di Intelligenza Artificiale, quindi ai tentativi di incorporare l'intelligenza umana nelle macchine e far effettuare loro determinate operazioni. Uno dei principali problemi legati all'utilizzo delle AI è quello etico della responsabilità umana, del *decision making*, quindi la delegazione di decisioni e compiti, di cui sicuramente si può discutere in un'ulteriore sede di analisi<sup>18</sup>.

Tra le tecnologie più utili per la Digital History e, più in generale, per le Digital Humanities, si annoverano il Machine Learning, il Deep Learning e il Text Mining.

Il Machine Learning (Apprendimento Automatico) è la capacità da parte di un sistema di utilizzare dati ed esperienze pregresse nello svolgimento di un compito con lo scopo di ottimizzare il processo operativo (Zinoviev 2016). La macchina riuscirà a svolgere nel tempo determinate funzioni, migliorando e sviluppando le proprie capacità con l'esperienza, mediante algoritmi che, partendo da nozioni primitive, sapranno prendere una specifica decisione piuttosto che un'altra o effettuare azioni apprese precedentemente. Il Machine Learning può essere supervisionato o non supervisionato: nel primo caso vengono fornite alla macchina nozioni etichettate (che vanno quindi a costituire un *database*) in modo che questa possa recuperarle e decidere quali utilizzare per rispondere a determinati compiti. Tipico esempio è il filtro *anti-spam* delle caselle di posta elettronica, che riconosce e segnala come *spam* alcune mail che, in base ad esperienze passate, potrebbero far parte di quella categoria (Muller and Guido 2016).

Nel secondo caso la macchina apprende informazioni senza che queste siano codificate, per cui dovrà catalogarle, organizzarle e comprendere così il loro significato.

Un'applicazione di Machine Learning in ambito storico-culturale è rappresentata dall'Automatic Analysis of Image Genres. Un team di studiosi, dopo aver digitalizzato con tecnologia OCR (Optical Character Recognition) milioni di pagine di giornali storici<sup>19</sup> e dopo aver processato e categorizzato gli articoli a livello testuale, ha analizzato i generi delle immagini attraverso l'Eigenvector Based Feature Modeling, suddividendole in 6 categorie (ritratti, parole, vista esterna degli edifici, mezzi busti, gruppi di persone) e utilizzando metodi di Machine Learning (Back Propagation e Simulated Annealing) per inserirle automaticamente all'interno delle categorie corrispondenti (Allen et al. 2010).

Il Text Mining invece riduce gli sforzi dello studioso nella ricerca con lo scopo di ottenere informazioni da una vasta quantità di testi analizzati, utilizzando il NLP (Natural Language Processing, elaborazione del linguaggio naturale).

"Text Mining can be defined as a knowledge-intensive process in which a user interacts with a collection of documents by using analytic tools in order to identify and explore interesting patterns. [...] Computer are able to analyze only the *syntactic* (enfasi di Žižka) aspect of texts, which means that they are able to recognize how words are arranged in the documents." (Žižka et al. 2020, 2).

Il NLP è il trattamento informatico (*computer processing*) del linguaggio naturale, quindi la lingua umana che contiene spesso e volentieri ambiguità e sottintesi<sup>20</sup>. L'applicazione del NLP permette la trasformazione del testo in dati strutturati e normalizzati, in modo tale da poter

<sup>18</sup> Si rimanda in primo luogo al testo a cura della Task force sull'Intelligenza artificiale dell'Agenzia per l'Italia Digitale, *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino*, versione 1.0 marzo 2019, disponibile su: <https://ia.italia.it/assets/librobianco.pdf>, oltre che al *White Paper on Artificial Intelligence* pubblicato il 19 febbraio 2020 dalla Commissione Europea di Bruxelles e disponibile su: [https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust\\_it](https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_it).

<sup>19</sup> National Digital Newspaper Program, About the program, <https://www.loc.gov/ndnp/about.html/>.

<sup>20</sup> Per un approfondimento si veda la traduzione di un articolo del Dr. Wei Li, ricercatore dell'azienda NetBase Solutions Inc., *Che cos'è il Natural Language Processing*, in <https://www.celi.it/blog/2015/10/che-cose-il-natural-language-processing/>

individuare i gruppi tematici principali, classificare i documenti in categorie, scoprire associazioni nascoste ed estrarre informazioni specifiche (Žižka et al. 2020).

Un esempio di applicazione di Text Mining nella ricerca storica è stato portato avanti da un gruppo di ricercatori che ha analizzato bollettini prodotti dal Comando Supremo del Regio Esercito Italiano nelle due guerre mondiali attraverso l’NLP e l’Information Extraction (Boschetti et al. 2014). Il gruppo ha svolto un’analisi statistica dei testi e ha estratto i più utilizzati termini di guerra, nomi propri e tipologie di eventi (bombardamenti, battaglie, etc.), collegando attraverso *link* diverse tipologie di dati a fonti esterne e sviluppando un’interfaccia di ricerca che ne permette l’esplorazione.

La più recente implementazione del Machine Learning è il Deep Learning, quindi un apprendimento approfondito attraverso il quale la macchina sfrutta la ANN (Artificial Neural Network, rete neurale artificiale). Neurale va inteso nel senso biologico del termine: si tratta di algoritmi che agiscono su neuroni artificiali, i quali ricevono segnali, li integrano e li propagano. In questo caso le reti sono composte da almeno 3 strati: un *input layer* (che riceve i dati), un *hidden layer* (che li elabora) e un *output layer* (che contiene il risultato finale)<sup>21</sup>. A un livello più profondo si trovano le DNNs (Deep Neural Networks), utilizzate soprattutto per la visione artificiale e per l’applicazione del NLP: in storia dell’arte, ad esempio, il sistema riesce a identificare in un’opera lo stile di pittura, il genere e l’artista, mentre in archeologia aiuta alla classificazione delle sottosuperfici e alla scoperta di siti sepolti (Fiorucci et al. 2020).

Con queste modalità, dunque, i dati sono prodotti dalle macchine, ma queste devono essere affiancate dall’interpretazione da parte dello storico, che deve conservare un adeguato margine interpretativo e la capacità critica di analisi e di confronto.

La problematica principale è rappresentata dall’affidabilità delle macchine e del dato. La macchina può tralasciare o travisare qualche informazione che invece è per lo storico di una certa importanza, per un motivo intrinseco nella sua natura: non può pensare come lo storico. Per quanto la tecnologia possa progredire, bisogna tener conto di questo limite, e di eventuali *bias*, “pregiudizi” insiti nelle procedure algoritmiche delle AI che possono causare discriminazioni e dar vita a disuguaglianze sociali e non solo. Basti pensare al COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), un particolare algoritmo utilizzato nei tribunali statunitensi in grado di predire il rischio di recidività di un imputato. Da un’analisi di ProPublica si evince che questo algoritmo presenta dei pregiudizi infondati contro chi ha la pelle scura (45% di probabilità di commettere reati) rispetto alla controparte bianca (23%)<sup>22</sup>.

Nelle procedure di Machine Learning, o più in generale nella creazione di dati (Data Generation), i *bias* possono essere di diversa tipologia, così come proposti recentemente in una tassonomia da Hellström, Dignum e Bensch. In questo caso una delle problematicità per uno storico, soprattutto nell’ambito della Text Analysis, può essere ritrovata negli *Inherited bias*: quando alcuni strumenti di Machine Learning che presentano già dei *bias* vengono usati per dar vita a input per altri algoritmi di Machine Learning (per esempio il caso del NLP), possono trasmettere ereditariamente i *bias* ai *tools*. I *bias* possono essere causati in questa situazione da vari procedimenti, come ad esempio quello della *sentiment analysis*<sup>23</sup>, della *word embedding*<sup>24</sup> o della più basilare *machine translation*, la traduzione automatica (Hellström et al. 2020).

<sup>21</sup> Per un approfondimento si veda Spindox, Machine Learning. Reti neurali demistificate, <https://www.spindox.it/blog/ml1-reti-neurali-demistificate/>.

<sup>22</sup> Per l’analisi si legga Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin. “How We Analyzed the COMPAS Recidivism Algorithm”, 23 maggio 2016, <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. La storia dell’inchiesta invece è disponibile su <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

<sup>23</sup> “La Sentiment analysis, anche detta *opinion mining*, è un ambito di studi che analizza l’opinione delle persone, i sentimenti [...] e le loro emozioni attraverso entità e attributi espressi a livello testuale.” Traduzione dell’autore da Liu, Bing. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge: Cambridge University Press, 2015: 1.

<sup>24</sup> Il *word embedding* è un insieme di tecniche di modellazione del linguaggio in cui parole o frasi vengono mappate in vettori di numeri reali. Grazie al *word embedding* è possibile cogliere il contesto delle parole in un documento, le somiglianze semantiche e sintattiche, relazioni con altre parole, etc. Si veda Dhruvil Karani. “Introduction to Word Embedding and Word2Vec”, in <https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa>.

## Presentazione delle esperienze di caso: applicazioni della Digital History

### Slave Voyages: Database non testuali e Data Visualization

Il progetto Slave Voyages è presente sul Web dal 2008 per opera della Emory University ma è frutto di un'attività di ricerca e raccolta di dati che getta le sue basi alla fine degli anni '60 nella persona di Herbert S. Klein e di altri studiosi che collezionano dati archivistici sulla tratta degli schiavi nell'ottica di renderli comprensibili alle macchine. I vari *databases* vengono raccolti negli anni dagli studiosi, soprattutto grazie all'incontro di David Eltis e Stephen Behrendt, che nel 1990 immaginano un *multisource database* che possa standardizzare i dati esistenti, collezionarli e aggiungere a questi nuove informazioni. Grazie all'aiuto di fondazioni, istituti e università presenti anche in altri Paesi, la documentazione si accresce numericamente fino a raggiungere i quattro continenti coinvolti negli oltre trentamila viaggi raccolti (Europa, Africa, Nord e Sud America)<sup>25</sup>.

L'interfaccia si presenta nella sua *Homepage* con un'animazione nello sfondo e diversi contenuti proposti, tra i quali un breve video introduttivo nel quale Henry Louis Gates Jr., della Harvard University, illustra la storia del progetto. Questo si divide principalmente in tre filoni: *Trans-Atlantic*, *Intra-American* e *African Names*. Il primo *database* riguarda la rotta atlantica, il secondo la rotta tra le due Americhe e il terzo le origini africane degli schiavi (nomi, regioni, etc.).

Ogni progetto offre contenuti di Data Visualization differenti: in questa sede risulta più utile analizzare esclusivamente il primo dei tre, lasciando al lettore l'esplorazione autonoma del secondo e del terzo, strutturati in una maniera quasi del tutto analoga.

*Trans-Atlantic* è il contenitore più ricco e offre diverse possibilità di interrogazione dei dati. Si possono infatti visualizzare i dati uno per uno, con l'opzione *Result*, ordinandoli per ID, nome del vascello, porto di partenza, principale luogo d'acquisto degli schiavi, porto di arrivo, nome e cognome del capitano, data di arrivo e numero di schiavi sbarcati. Inoltre, con *Summary statistics* e *tables* si possono formare delle tabelle con dei resoconti più generali, mentre la Data Visualization permette di rappresentare i dati all'interno dei grafici cartesiani, scegliendo le variabili nell'asse delle ascisse e delle ordinate.

Con la *Timeline* si può visualizzare una classica linea del tempo, selezionando ancora una volta una variabile da considerare. Con *Maps* è possibile visualizzare una cartina geografica interattiva che presenta le rotte e le destinazioni dei vascelli con specifiche informazioni a riguardo; con il *Timelapse* si assiste a una visualizzazione che ripercorre dal 1519 al 1860 tutte le rotte con delle animazioni che si muovono da una parte all'altra del globo a rappresentare le navi che ripercorrono le tratte storicamente compiute<sup>26</sup>. Una sezione interna al *Trans-Atlantic* è dedicata ai saggi di approfondimento (*Essays*) realizzati dai ricercatori e liberamente accessibili, mentre un'altra categoria è dedicata alla spiegazione della metodologia di lavoro che è stata seguita per realizzare il *database*<sup>27</sup>; infine la sezione *Estimates* raccoglie delle stime numeriche riguardanti diversi argomenti. Tutti i dati sono scaricabili, *open* e possono essere interrogati mediante *query*. Un ulteriore contenuto multimediale offerto da questa sezione è rappresentato da un breve video in cui viene presentata una ricostruzione in versione 3D di un vascello del 1800, l'*Aurora*<sup>28</sup>. È inoltre presente una ricca galleria di immagini corredate da *metadata* e descrizioni, nella quale spiccano manoscritti originali e luoghi, oltre a raffigurazioni di schiavi e vascelli<sup>29</sup>.

### Alcide Digitale: Text Analysis e Natural Language Processing

Alcide Digitale è un progetto creato e sviluppato nel 2019 dalla Fondazione Bruno Kessler di Trento in collaborazione con l'Istituto Storico Italo-Germanico e con il contributo di Fondazione Cariplo, Fondazione Caritro e della casa editrice il Mulino. Alcide, oltre ad essere il nome dello statista italiano De Gasperi, è anche l'acronimo di Analysis of Language and Content In a Digital

<sup>25</sup> Slave Voyages, About, <https://www.slavevoyages.org/about/about#>.

<sup>26</sup> Trans-Atlantic Slave, Database, <https://www.slavevoyages.org/voyage/database#timelapse/>.

<sup>27</sup> Trans-Atlantic Slave, Understanding The Database, <https://www.slavevoyages.org/voyage/about#methodology/introduction/0/en/>.

<sup>28</sup> Trans-Atlantic, Slave Ship in 3D Video, <https://www.slavevoyages.org/voyage/ship#slave->

<sup>29</sup> Trans-Atlantic, Image Galleries, <https://www.slavevoyages.org/resources/images/>.

Enviroment: il progetto si prefigge lo scopo di analizzare i documenti prodotti da De Gasperi<sup>30</sup>, permettendo all'utente l'estrazione di concetti chiave, la ricerca testuale, il riconoscimento di nomi di persona e di luogo, la visualizzazione dei network, l'individuazione delle co-occorrenze, l'analisi della complessità sintattica e semantica dei testi<sup>31</sup>. Alcide è il risultato dell'applicazione di procedure di Text Mining e più in generale di Text Analysis, quindi sono stati utilizzati *tools* per la segmentazione del testo, per l'estrazione di lemmi e di parti del discorso, nomi di persona, luoghi geografici e concetti chiave; i dati ottenuti sono poi presentati tramite Data Visualization grazie ad alcune estensioni dell'applicazione ReactJS<sup>32</sup>.

Prima di accedere direttamente alla piattaforma ALCIDE, nella sua *Homepage* il sito offre all'utente vari spunti per un'esplorazione guidata. Una prima soluzione è data dalla *timeline* che scandisce la vita e l'opera di De Gasperi in cinque periodi, poi ulteriormente suddivisi in base al ruolo da lui svolto. Scegliendo uno di questi percorsi si può accedere alla piattaforma, quindi si possono analizzare documenti, relazioni, nomi di persone e luoghi, temi, etc... Un'altra soluzione di esplorazione è offerta da una parte di pagina dedicata a uno *storytelling* che introduce l'utente a documenti o ambiti ben precisi, definiti a livello tematico.

Accedendo alla piattaforma si visualizza un'interfaccia interattiva con cui effettuare la ricerca secondo varie possibilità: *Ricerca testuale*, *Rete persone*, *Luoghi*, *Persone*, *Parole chiave*, *Co-occorrenze*.

La funzione *Cerca nel testo* permette di trovare specifici lemmi o stringhe di parole all'interno del testo; inoltre fornisce un *n-gram* con la frequenza con la quale la parola o le parole ricercate si ritrovano nel *database*. Con lo strumento *Tempo e contesto* è possibile selezionare l'arco di tempo desiderato e la tipologia di fonte documentaria. Al fondo di ogni testo è presente un doppio grafico che segnala sia la presenza di eventuali nomi nel testo sia le parole chiave che riassumono l'argomento; un brevissimo *abstract* fa da sottotitolo a ogni documento.

Un altro *tool* presenta un diagramma con le parole chiave più frequenti nel *database*, permettendo la ricerca anche secondo questo criterio. Con *Co-occorrenze* è possibile estrarre dei termini che co-occorrono con una chiave di ricerca, quindi si ha la possibilità di visualizzare i documenti che fanno riferimento a quella parola che a sua volta può essere associata a un'altra. Con la ricerca *Persone* è possibile ritrovare nomi e cognomi di persone presenti nei testi, mentre con la ricerca *Luoghi* in una mappa cartografica interattiva si possono ritrovare sia i luoghi di pubblicazione dei documenti che i luoghi menzionati all'interno degli scritti.

Ultimo *tool* offerto dalla piattaforma è *Rete Persone*, che permette di visualizzare una "rete di persone" connesse tra loro grazie a un termine che l'utente può selezionare. Anche in questo caso i risultati rimandano ai testi digitalizzati e presenti nel *database*, punto terminale della piattaforma.

## Discussione critica dei casi esaminati

Nel *mare magnum* di progetti afferenti all'area delle Digital Humanities, e, più in particolare, della Digital History, i due casi analizzati rispondono ad alcune delle più importanti esigenze di ricerca degli storici moderni ma soprattutto contemporaneisti, cioè l'analisi che si svolge all'interno di un *database* di fonti e la sua visualizzazione. I due progetti, pur avendo una comune natura di fondo, si differenziano *in primis* per l'area territoriale: *Slaves Voyages* è per ovvi motivi di interesse globale, mentre ALCIDE spazia su un'area definita dalla posizione del noto statista, perlopiù Europa Occidentale e Stati Uniti Orientali. Se il primo è un database quantitativo più che qualitativo, il secondo rappresenta l'esatto opposto, trattandosi di un database testuale. In definitiva i diversi *tools* e le possibilità di ricerca offrono tanti spunti di riflessione che non si limitano semplicemente al ramo della ricerca storica ma si allargano verso orizzonti multidisciplinari.

*Slaves Voyage* è quasi sicuramente il più fresco, interattivo e ricco lavoro che utilizza la Data Visualization in ambito prettamente storico, facendo riferimento a un periodo che potrebbe essere definito, per riprendere un concetto caro a Braudel e alla scuola francese degli *Annales*,

<sup>30</sup> Riprendendo come corpus di fonti l'intera edizione A. De Gasperi, *Scritti e discorsi politici*, I-IV, sotto la direzione scientifica di Paolo Pombeni, Bologna, il Mulino, 2006-2009.

<sup>31</sup> Alcide, Un viaggio negli scritti di De Gasperi, *Homepage*, <http://alcidedigitale.fbk.eu/>.

<sup>32</sup> React – A JavaScript library for building user interfaces, <https://reactjs.org/>.

di *Longue Durée*<sup>33</sup>. Il database si concentra quindi su una struttura storica di lunga durata come quella della tratta degli schiavi, considerata a livello documentario in questo caso dal 1514 al 1866. La piattaforma si focalizza molto più sull'analisi quantitativa che su quella qualitativa, in una visione molto vicina a quella delle scienze sociali. In questo senso può essere d'aiuto per approfondimenti e ricerche statistiche sulla gigantesca questione della tratta degli schiavi, fornendo a storici, studiosi e appassionati numeri e stime a riguardo. Allo stesso tempo offre diverse possibilità di esplorazione e soprattutto la prospettiva dell'*open data*, che, a parere di chi scrive, non può essere assolutamente ignorata nelle ricerche scientifiche del XXI secolo, tanto più se un lavoro del genere è sostenuto da molte realtà (perlopiù universitarie) presenti sul territorio americano, una tra tutte l'Hutchins Center della Harvard University. Il *database* è quindi molto ricco, frutto sì di tanti anni di lavoro ma anche di un quotidiano aggiornamento da parte di un team altamente specializzato. Sul versante più *visual* e grafico la sua interfaccia è dinamica, semplice e adatta a qualsiasi tipo di pubblico. Lo *storytelling* è funzionale e le possibilità di approfondimento sono di vari livelli, dalla più superficiale alla più profonda. Tutto sembra essere molto scorrevole e non si evidenziano particolari limiti tecnici; l'utente può inoltre segnalare eventuali dati non corretti o mancanze e contribuire con fonti e note di propria conoscenza.

La piattaforma AlcideDigitale, infine, è realizzata per qualsiasi utente voglia approfondire l'opera degasperiana. A tal proposito Maurizio Cau, che ha supportato il progetto sul versante storico, lo definisce come "Un primo esperimento intenzionato a fornire contenuti per un pubblico generalista e offrire strumenti di analisi non ordinaria per gli storici."<sup>34</sup> Il progetto si presta molto ai vari livelli di approfondimento e di ricerca, con un'interfaccia *user-friendly* in cui tutto appare ordinato e schematico, presentato in una *Homepage* statica a discesa. Tuttavia Alcide è un po' carente per quanto riguarda la parte di *storytelling*, che si limita a una breve introduzione degli argomenti, senza presentare più approfonditamente il personaggio, permettendo comunque all'utente di muoversi con parecchia libertà esplorativa. Un altro limite del lavoro risiede nella sua bibliografia: tutto il materiale proviene unicamente dai quattro volumi che raccolgono gli scritti e i discorsi politici di De Gasperi pubblicati dalla casa editrice il Mulino, per cui Alcide non è un *database* che raccoglie tutta l'opera dello statista; allo stesso tempo non si possono effettuare comparazioni all'esterno del contesto degasperiano.

"La piattaforma permette l'osservazione a distanza, ma è una forte interpolazione affidata alla macchina; serve sempre l'intervento interpretativo dello storico, perché la macchina non esaurisce mai l'orizzonte contestuale entro cui ci si muove."<sup>35</sup>

Nonostante questi limiti il progetto rappresenta un innovativo tentativo di collaborazione tra umanisti e informatici, in quanto permette una più che dettagliata ricerca testuale e soprattutto *Open Access* di buona parte di testi che dal punto di vista storico assumo un particolare valore. Allo stesso tempo si tratta l'applicazione mostra alcune delle più efficaci possibilità di utilizzo delle AI nella ricerca storica e filologica, quindi il Text Mining e il Natural Language Processing. L'analisi e la ricerca testuale, o ancora il poter relazionare documenti, temi, luoghi e persone, offrono al visitatore uno sguardo più ampio sulla produzione degasperiana e sul contesto che ruota intorno a questa figura. Come sottolineato da Cau, la varietà di possibilità di osservazione può generare dei punti di vista molto interessanti per uno storico o un filologo, che grazie ai vari *tools* può scovare relazioni e collegamenti che prima probabilmente non avrebbe notato con la stessa facilità; a ciò si aggiunge una piacevole interfaccia grafica che trascina la produzione letteraria in un contesto digitale per certi versi accattivante.

## Conclusioni

Dal punto di vista metodologico, quando si fa riferimento alla Digital History, non si può parlare di una rivoluzione scientifica, in quanto questa prevede la sostituzione di paradigmi passati con

<sup>33</sup> Per un approfondimento si veda Braudel, Fernand, and Immanuel Wallerstein. "History and the Social Sciences: The Longue Durée." *Review (Fernand Braudel Center)* 32, no. 2 (2009): 171-203. Accessed December 17, 2020. <http://www.jstor.org/stable/40647704>.

<sup>34</sup> Intervista dell'autore a Maurizio Cau, ricercatore presso la Fondazione Bruno Kessler, in data 25/05/2020.

<sup>35</sup> Intervista dell'autore a Maurizio Cau, ricercatore presso la Fondazione Bruno Kessler, in data 25/05/2020.

altri nuovi, quindi il cambiamento di regole, canoni e metodi (Kuhn, 2009). La ricerca storica e la sua euristica partono da una o più domande e, attraverso una metodologia lavorano per trovare delle soluzioni, delle risposte. In questo caso non si tratta ancora di una crisi della disciplina in quanto tale e dei suoi paradigmi, ma sicuramente questo nuovo ambiente offre spazio a momenti di riflessione e di discussione sulla legittimità di alcuni metodi, problematiche e modelli di soluzione.

Da qui l'evidente difficoltà nella definizione di Digital History: se "History" è la disciplina storica, il significato dell'aggettivo "Digital" risulta quanto mai complesso, perché il digitale ormai è entrato in tutte le attività quotidiane, dal lavoro alle relazioni, dalla burocrazia agli studi e alle ricerche. La condizione di *lockdown*, dovuta alla diffusione del *Covid-19*, all'interno della quale si sono trovati e si trovano diversi Stati ha significativamente dimostrato l'importanza del digitale nell'età contemporanea. Per citare Massimo Durante, che fa riferimento alla rivoluzione digitale ampiamente trattata da Floridi con tutte le sue conseguenze e influenze:

"La rivoluzione, basata sul potere computazionale, è soprattutto una rivoluzione del quotidiano: tanto più profonda e diffusa quanto più investe la trama delle nostre abitudini e pratiche consuete, cioè, le nostre forme di vita." (Durante 2020, 11)

Di conseguenza nemmeno la Digital History può essere definita una sottodisciplina della Storia, in quanto non presenta un insieme definito di regole, canoni o metodi peculiari e specifici. In questo senso può essere momentaneamente intesa come un approccio, o un insieme di pratiche (continuamente in evoluzione), o uno strumento di grande supporto al lavoro dello storico, il quale può trarne in diversi sensi utilità nelle varie applicazioni. Allo stesso tempo la ricerca deve porre attenzione ai nuovi spunti della disciplina offerti dall'influenza e dal persistente progresso della tecnologia in modo da poterli analizzare con consapevolezza critica. È necessario sottolineare che l'ambiente della Digital History non fa riferimento esclusivamente alla periodizzazione della storia contemporanea, ma include – nelle sue diverse coniugazioni – tutti gli ambiti di ricerca storica, dal meno recente in avanti, ognuno con le sue peculiarità e le possibilità di approfondimento in ambiente digitale: analisi diplomatica e metadattazione di *corpus* di fonti medioevali, archivistica moderna, Historical Gis per la geografia storica, solo per riprendere alcuni esempi. In questo senso la Digital History può fornire spunti di approfondimento e di studio mirati ad un'analisi e all'applicazioni di modelli qualitativi, quali lo studio e l'interpretazione del documento o della fonte.

In una prospettiva più ampia, la via che si sta seguendo punta a una facilitazione del lavoro di ricerca, soprattutto dal punto di vista testuale: in tal modo lo storico riesce a lavorare su vasti orizzonti in un minor lasso di tempo. Gli strumenti a disposizione del Digital Historian, figli dell'applicazione del potere computazionale, hanno creato un *gap* rispetto alle tradizionali procedure in termini di accessibilità, rapidità, efficienza e semplificazione del lavoro. Lo storico può spesso lavorare online, direttamente dal proprio computer, avendo a disposizione grandi banche dati (testuali e non) da esaminare con *tools* appositi (Data e Text Mining), fonti digitalizzate da visualizzare in alta definizione (tecnologie OCR), documenti da catalogare e ordinare (metadati e linguaggi di marcatura) in tempi mai visti prima. Tuttavia alla semplificazione non deve e non può corrispondere da parte dello storico un'attribuzione delle responsabilità alla macchina, i cui procedimenti e risultati devono essere sottoposti a una verifica da parte del primo, affiancato da un *team* di professionisti che possa colmare le sue eventuali lacune tecnico-informatiche. Come inoltre si è visto, non è da trascurare il fronte dei rischi, delle possibilità di errore e dei *bias* presenti in alcune procedure di Machine Learning e in generale nelle AI. Un semplice *bias* ereditato, un valore o un dato inesatto, potrebbe alterare la funzionalità e l'efficacia degli algoritmi, oltre che la natura stessa dei dati, delle fonti e conseguentemente dei risultati che la macchina raggiunge e rilascia.

Il caso di ALCIDE presenta un prototipo di ricerca che può assolutamente essere applicato anche in altri contesti, soprattutto laddove la filologia e l'elemento testuale assumono un peso rilevante, quindi ad esempio gli studi di storia del pensiero politico o storia delle religioni. Allo stesso tempo i database non testuali, sulla rotta tracciata da *Slaves Voyages*, risultano essere utili in una visione più statistica e quantitativa della storia e di altri ambiti disciplinari contigui come quello della storia politica, militare o economica. Il modello di Data Visualization può supportare l'attività dello storico, perlomeno sul piano della visualizzazione dei contenuti, e configurarsi come un chiaro strumento di presentazione di dati, ricerche e risultati. Inoltre, essendo il digitale uno spazio aperto a tutti, lo storico deve valutare e porre particolare attenzione alla dimensione pubblica, quindi la Digital Public History, creando percorsi quanto più

accessibili. Ciò non significa che tutta la storia debba diventare Digital Public History, né tantomeno Digital History: lo storico deve essere in grado di distinguere i vari livelli di comunicazione che, a parere di chi scrive, sono destinati a coesistere e devono in questo senso poter dialogare e supportarsi a vicenda.

In conclusione, si può senza grosse difficoltà sostenere l'integrazione tra i diversi approcci come soluzione provvisoria sulla riflessione teorico-metodologica dell'ambiente della ricerca storica. Se la disciplina tradizionalmente intesa infatti offre alcuni imprescindibili strumenti di lavoro oltre che il rigore scientifico, la Digital History amplia le possibilità di studio, di approfondimento e di accessibilità, fornendo allo stesso tempo nuovi strumenti e nuovi punti di vista in grado di aiutare lo studioso, oltre che spunti di riflessione metodologica che non possono più essere ignorati.

## Bibliografia

- Allen, Robert. B., Weizhong Zhu, and Robert Sieczkiewicz. *What to Do With a Million Pages of Digitized Historical Newspapers?* (2010). Accessed December 17, 2020. <https://www.semanticscholar.org/paper/What-to-Do-With-a-Million-Pages-of-Digitized-Allen-Zhu/d9c4c54ed8c1f06e8a809ac8ef5b56830048a5f8>
- Bloch, Marc. *Apologia della storia o Mestiere di storico*. Torino: Einaudi, 2009 (Apologie pour l'histoire ou Métier d'historien, 1993).
- Boschetti, Federico, Andrea Cimino, Felice Dell'Orletta, Gianluca Lebani, Lucia Passaro, Paolo Picchi, Giulia Venturi, Simonetta Montemagni, and Alessandro Lenci. *Computational Analysis of Historical Documents: An Application to Italian War Bulletins in World War I and II*. Lussemburgo: ELRA, 2014: 70-75. Accessed December 17, 2020. <https://arpi.unipi.it/retrieve/handle/11568/686685/80767/WW-LREC2014.pdf>
- Bozic, Bojan, Gavin Mendel-Gleason, Christophe Debruyne, and Declan O'Sullivan. *Computational History and Data-Driven Humanities, Second IFIP WG 12.7 International Workshop CHDDH 2016, Dublin, Ireland, May 25, 2016, Revised Selected Papers*. Cham: Springer, 2016.
- Castells, Manuel. *Comunicazione e potere*. Milano: Università Bocconi, 2014 (Communication Power, 2009).
- Cauvin, Thomas. *Public History: A Textbook of Practice*. New York: Routledge, 2016.
- Cellini, Paolo. *La rivoluzione digitale*. Roma: LUISS, 2018.
- Detti, Tommaso, and Giuseppe Lauricella. "Una storia piatta? Il digitale, Internet e il mestiere di storico." *Contemporanea*, 10.1 (2007): 3-23. Accessed December 17, 2020. <https://www.jstor.org/stable/24653055>.
- Delle Donne, Roberto. "Le fonti digitali e le pratiche della ricerca. A proposito di un recente libro di Stefano Vitali." *Reti Medievali*, 6.2 (2005): 1-17. Print.
- Durante, Massimo. *Potere computazionale. L'impatto delle ICT su diritto, società, sapere*. Milano: Meltemi, 2020.
- Falletta, Serena. "I linguaggi di marcatura nel trattamento informatico delle fonti storiche: l'edizione digitale del codice Vat.Lat.3800." *Schede Medievali*, 48 (2010): 253-271. Print.
- Fiorucci, Marco, Marina Khoroshiltseva, Massimiliano Pontil, Arianna Traviglia, Alessio Del Bue, and Stuart James. "Machine Learning for Cultural Heritage: A Survey." *Pattern Recognition Letters* 133 (2020): 102-108. Accessed December 17, 2020. <https://doi.org/10.1016/j.patrec.2020.02.017>

- Floridi, Luciano. *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Milano: Raffaello Cortina, 2017 (The Fourth Revolution. How the Infosphere is Reshaping Human Reality, 2014).
- François, Pieter, Manning J.G., Harvey Whitehouse, Rob Brennan, Thomas Currie, Kevin Feeney, and Peter Turchin. "A Macroscope for Global History: Seshat Global History Databank, a methodological overview". *Digital Humanities Quarterly*, 10.4 (2016). Accessed December 17, 2020. <http://www.digitalhumanities.org/dhq/vol/10/4/000272/000272.html>.
- Ghos, Ashish. *Multi-objective evolutionary algorithms for knowledge discovery from databases*. London: Springer, 2008.
- Grandi, Elisa, and Émilien Ruiz. "Ce que le numérique fait à l'historienne. Entretien avec Claire Lemercier." *Diacronie* 2.10 (2012). Accessed December 17, 2020. <https://doi.org/10.4000/diacronie.2780>
- Guerrini, Mauro, and Tiziana Possemato. *Linked data per biblioteche, archivi e musei*. Milano: Editrice Bibliografica, 2015.
- Hellström, Thomas, Virginia Dignum, and Suna Bensch. "Bias in Machine Learning – What is it Good for?". *arXiv*, 20 settembre 2020. Accessed December 17, 2020. <https://arxiv.org/pdf/2004.00686.pdf>.
- Henrickson, Leah. "Humanities Computing, Digital Humanities and Computational Humanities: What's in a Name" – *3:AM Magazine*. Accessed December 17, 2020. <https://www.3ammagazine.com/3am/humanities-computing-digital-humanities-and-computational-humanities-whats-in-a-name/>
- Hurwitz, Judith, Alan Nugent, Fern Halper, and Marcia Kaufman. *Big Data for Dummies*. Hoboken: Wiley, 2013.
- Kirk, Andy. *Data Visualisation, A Handbook for Data Driven Design*. London: SAGE Publications Ltd, 2016.
- Kuhn, Thomas S. *La struttura delle rivoluzioni scientifiche*. Torino: Einaudi, 2009 (The Structure of Scientific Revolutions, 1962).
- Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. Boston: Houghton Mifflin Harcourt, 2013.
- Muller, Andreas C., and Sarah Guido. *Introduction to Machine Learning with Python. A Guide for Data Scientists*. Sebastopol: O'Reilly, 2016.
- Nelson, Lynn H. "Before the Web: The Early Development of History On-line", edited by Serge Noiret, Dossier su *Internet ed il Mestiere di Storico*, a cura di Serge Noiret. SISSCO, Società italiana per lo studio della storia contemporanea. Accessed December 17, 2020. <http://chnm.gmu.edu/digitalhistory/links/pdf/chapter1/1.2b.pdf>
- Nilsson Hammar, Anna. "Digital History." *Scandia introducerar* 81.2 (2015): 99-110. Accessed December 17, 2020. <https://journals.lub.lu.se/scandia/article/view/17287>
- Noiret, Serge. "Informatica, storia e storiografia: la storia si fa digitale." *Memoria e Ricerca: rivista di storia contemporanea* 28 (2008), pp. 189-201. Accessed December 17, 2020. doi:10.1400/115117.
- Noiret, Serge. "Storia Digitale. Quali sono le risorse di rete usate dagli storici?", in *La macchina nel tempo. Studi di informatica umanistica in onore di Tito Orlandi*, edited by Lorenzo Perilli, Domenico Fiormonte, 201-258. Firenze: Le Lettere, 2011.

- Numerico, Teresa, Domenico Fioramonte, and Francesca Tomasi. *L'umanista digitale*. Bologna: Il Mulino. 2010.
- Nyhan, Julianne, and Andrew Finn. *Computation and the Humanities. Towards an Oral History of Digital Humanities*. London: SpringerOpen. Accessed December 17, 2020. DOI 10.1007/978-3-319-20170-2
- Nystrom, Elsa A., and Justin A Nystrom. "Beyond the Valley of the Shadow: Taking Stock of the Virginia Center for Digital History." *Merlot, Journal of Online Learning and Teaching* 1.1 (2005). Accessed December 17, 2020. [https://jolt.merlot.org/vol1\\_no1\\_nystrom.htm](https://jolt.merlot.org/vol1_no1_nystrom.htm)
- Pancur, Andrej, and Mojca Sorn. "Smart Big Data: Use of Slovenian Parliamentary Papers in Digital History.", *Prispevki za Novejšo Zgodovino*, 56.1 (2016): 130-146. Accessed December, 17, 2020. <https://ojs.inz.si/pnz/article/view/193/296>
- Salvatori, Enrica. "Digital (Public) History: la nuova strada di una antica disciplina." *RiMe. Rivista dell'Istituto di Storia dell'Europa Mediterranea* 1.1 (2017): 57-94. Accessed December 17, 2020. <https://doi.org/10.7410/1291>
- Seefeldt, Douglas and William Thomas. "What is Digital History?" *Perspectives on History*, May 1, 2009. Accessed December 17, 2020. <https://www.historians.org/publications-and-directories/perspectives-on-history/may-2009/what-is-digital-history>
- Thaller, Manfred. "The Need for a Theory of Historical Computing", in *History and Computing II*, ed. Peter Denley et al., 2-11. Manchester: Manchester University Press, 1991.
- Tomasi, Francesca. "Digital Humanities e organizzazione della conoscenza: una pratica di insegnamento nel LODLAM", *AIB Studi* 60.2 (2020). Accessed December 17, 2020. <https://aibstudi.aib.it/article/view/12068>.
- Van Eijnatten, Joris, Toine Pieters, and Verheul, Jaap. "Big Data for Global History. The Transformative Promise of Digital Humanities." *BMGN – Low countries Historical Review* 128.4 (2013): 55-77. Print.
- Vitali, Stefano. *Passato digitale. Le fonti dello storico nell'era del computer*. Milano: Bruno Mondadori, 2004.
- Vivarelli, Maurizio. "Digital Humanities e culture documentarie: un modello di analisi, valutazione e interpretazione.", *AIB Studi*, 60.3 (2020). Accessed February, 26, 2021. <http://dx.doi.org/10.2426/aibstudi-12471>.
- Zaagsma, Gerben. "On Digital History." *BMGN – Low Countries Historical Review* 128.4 (2013): 3-29. Print.
- Zinoviev, Dmitry. *Data Science con Phytton. Dalle stringhe al machine learning, le tecniche essenziali per lavorare sui dati*. Milano: Apogeo, 2016.
- Žižka Jan, Frantisek Darena, and Svoboda Arnost. *Text Mining with Machine Learning. Principles and Techniques*. Boca Raton: CRC Press, Taylor & Francis Group, 2020.